5-29-98, rev. 6-6-98, 6-24-98, 8-13-98, 2/11/99, 3-22-99, 6-11-99, 9-25-00, 8-23-01, 10-12-01

Flow of General Imputation

programs.gimp.rungimp.source will run the general imputation programs

HOWEVER...
*Processes which must run first:
(may vary somewhat between surveys and will be placed in a script)
-prep of the prior stat period
<SPLIT> WITH CURRENT STAT PERIOD SPLIT PERCENTAGES, DERIVED, ADJUST
*prep of the current stat period
SIMPLE IMPUTATION PCFLG<>'C',< SPLIT>, SIMPLE IMPUTATION PCFLG = 'C',DERIVED, ADJUST

Set response code on the C1 stat period control file:
1) Run a program to update the RSPCDE field on the stat period control file

Prepare the Fat Record:
1) Clear EFLG3 value from DATA00.ITyyyypp. This is related to H-B edit operations.

2) For surveys that choose to do so, reset imputed data to previously imputed values. This is specified in PARMLIB.SURVPARM (where parmname='GIMPRESET' and cvalue='Y'). Only data items where DATAFLAG in('F', 'M') and IMPACT ne 'N' and BYGIMP ne 'B' will be reset. The following variables will be reset: EDDATA, ADDATA, WDDATA, DATAFLAG, IMPFLAG, EFLG1-12 (except EFLG3).

3) Call CNVT with following options:
IDs - all, ITEMs - all, VERSIONS - E, A,W and J, CNTRL- all, STATPERIODS - take imputation relative stat periods (from PARMLIB.GRELSTAT) and convert to 'real' stat periods based on the base stat period, SUBSET - status = 'A' AND (whatever is in a user defined subsetting clause found in PARMLIB.PSUBSET for the MAKEGFAT record if user subsetting clause is not blank and status = blank) else just check status = 'A'
Output file - DATA00.GFAT
NOTE: late '99, we added capability of restricting the data versions on the fat record. If the file PARMLIB.DATAVERS has entries for module=IMPUTE, then only those versions will be fed to the CNVT routine and placed on the fat record.

4) Query the GIS file and make a temp file of those recodes that are necessary. Query to grelstatps file to find the needed relative stat period. Use this to then run the recodes from the SPRLGLIB.RCSOURCE catalog that are on the temp file to update the fat record.
Output file - updated DATA00.GFAT

Run H-B Edit

StEPS provides the ability to run H-B edit as part of a general imputation run. H-B edit failures will be indicated in the fat data set representation of IMPACT (2nd char of I<>00 and also J<>00) and EFLG3 (the 3rd char of G<>00). IMPACT will not be updated on item data sets. Once these fields are set, imputation will execute with the H-B edit failures appropriately dictating data imputes.

Valid values for the H-B flag (EFLG3 or 3rd char of G<>00) are as follows:

R - review
S - exclude from imputation base but impute only if edit reject
I - exclude from imputation base and force the item to be imputed
blank - no effect

These values translate into IMPACT (2nd char of I<>00 and J<>00) as follows:

H-B flag 'S' results in IMPACT='X'
H-B flag 'I' results in IMPACT='Y'

Create Imputation Reject File (aka step 1):

1) Ignore IDs where the BYGIMP = 'B'

*For records placed on the imputation reject file below- drop the prefix of the item names before placing on the output file.

2) Look at dataflag and at IMPFLG and place an ID/item on the imputation reject file if:

(dataflag = 'F' and impflg in (A,B,E,G,H,I,J,K,M,P,T,V,X)) or
(dataflag = 'M')
(NOTE: on 12/99 we use to instead have or (dataflag = 'M' and rspcde = 'I'))

4) Using edit1, edit2, edit3, edit6 (as long as the TESTSTAT= ' ' and g_event = 'Y' and the precondition is satisfied) determine which ID/items fail and place on the imputation reject file with IMP field set to 'Y'.

5) NOTE: we have commented out this part. May reinstitute as an option later if surveys need it.
Using the GIBS file (as long as TESTSTAT = ' ' and the precondition is satisfied) determine which balance complexes fail (ie. when sum of detail does not equal the total) and place all detail and total items on the imputation reject file with IMP field set to 'Y' or 'N' as determined by the IMPY and IMPXi fields on the GIBS file.

5') In 8/00 we reinstated something similar to 5 above:
Using the GIBS file (as long as TESTSTAT = ' ' and the precondition is satisfied) determine which balance complexes fail (ie. when sum of detail does not equal the total) and place all detail (not total items) on the imputation reject file with IMP field set to 'N'. If any of detail items or total item is imputed, then place all details on the imputation reject file with IMP field set to 'N').

6) Save the imputation reject file in DATA00.IMPREJ.

7) Unduplicate the imputation reject file by ID/item making sure to keep the record with IMP field of 'Y' if more than one ID/item record exist with different values of IMP.
Output file - DATA00.IMPREJUD

8) A program then takes this reject file and updates the GFAT fields J<item>## if the corresponding ID/item is on the reject file. A value of 'R' is placed in the respective field unless a value of 'Y' or 'N' is already there. (Y meant the impact field for the item was a Y, N meant impact field for the item was N)

9) This program will also add a new field to each ID record on the GFAT file called _COMPLEX by first using the balance complex parameter file, DATA00.GIBS and for each ID on the fat record determine if ANY of the balance complex records with TESTSTAT=' ' on the GIBS file are out of balance. If so, _COMPLEX is set to 'Y', else it is set to 'N'.

10) In 7/00 we added: At the end of IMPREJ.SOURCE, program now checks for the existence of $SPRGLIB/ckrecode.sas. If it exists, it is included and executed. This is designed specifically to clear the values of recoded variables if it is determined that these variables were computed using rejected items. Survey production programmers are responsible for developing and ensuring the accuracy of these programs.

Create Imputation Base (aka step 2):

1) (Pre-process the GIS file reading the global values and 'complete' the item and complex records where applicable. Also 'complete' GIS complex records when the method is IMPALL by obtaining necessary values from the corresponding GIBS record(s).
Output file - PARMLIB.GISGLBL)
NOTE: *presently we have just made a direct copy of GIS to GISGLBL until time to test pre-GIS.* No one is to date using the global feature and impall is no longer a valid method.

2) Ignore IDs on GFAT where: BYIMPB = 'B' or that do not meet the users' include clause as defined in file PARMLIB.PSUBSET for the CRE8IMPB record. (If the clause on the CRE8IMPB record is blank or the status is not blank, only ignore where BYIMPB = 'B').

3) There are 3 methods on the GIS file that require the imputation base: RATIO, AUXRAT, MEAN
For these methods determine ID/items that are to be used in calculating a ratio or mean and place numerator and denominator values on an intermediate file- DATA00.BASEDETL (NOTE: Numerator and denominator values specified by AUX variables can be the result of modifying the value of the AUX variable by use of a specified expression. See 6) below.)
(The BASEDETL data set is essentially a report of all ID's included in the imputation base by SPEC_NUM. As the imputation base program executes, an observation is written to the BASEDETL data set if it is deemed to meet the conditions for a method on the GIS data set. Note that an ID can appear more than once on BASEDETL because it may have met the criteria for more than one METHOD.)

(For year 1- we concentrated on RATIO and AUXRAT).

Identify the corresponding GIS record on this output file by the spec_num.
*it should include: ID, spec_num, nval, dval, catval, minnum, method, U, L, program date time

*Exclude following ID/ items:
- ID/items in which <u>any version</u> of the ID/item is on IMPREJUD;
- ID/items where impact <> ' ';
- ID/items that are missing,
- Numerator items (n) in which
(N1) the corresponding denominator value (d) is zero, missing, or <u>any version</u> of the denominator variable is on IMPREJUD,

(N2) the value of the ratio n/d > U, or
(N3) the value of the ratio n/d < L .

- Denominator items (d) in which
(D1) the denominator value is equal to 0,
(D2) the corresponding numerator value (n) is missing or <u>any version</u> of the numerator variable is on IMPREJUD, or
(D3) the value of the ratio n/d > U , or
(D4) the value of the ratio n/d < L

**in 6/99 we also decided that for RATIO and AUXRAT, if the data flag of either the numerator or the denominator was in (F,I,M,V,T), then that ID/ item should not be included in the imp base detail file and thus part of the imputation base.

METHOD=MEAN is not yet available, IMPALL is not longer a valid method.
(Notes: If METHOD=MEAN, then d=1.0 above. If METHOD=IMPALL and U=_NULL_, then disregard N2 and D3 above. If METHOD=IMPALL and L=_NULL_, then disregard N3 and D4 above.)

4) Using the intermediate file of ID/items for calculations, perform PROC SUMMARY and additional calculations as specified on the GIS and place on file - DATA00.BASESUM.

5) Then prepare the file for input to creating the final format. The START field (ie. match key) is presently the spec_num and the category combined and the VALUE field is usually the numerator / denominator for RATIO and AUXRAT methods. (If the number of IDs available to calculate an imputation quantity is less than MINNUM, set the imputation-base quantity equal to missing). This is placed on file-DATA00.BASFMTIN.

6) Using the summary format file, make the informat library which serves as the imputation base:
Output file - $IMPBASE

7) DON'T forget: AUX variables now also have additional fields to perhaps have more of a formula (eg. Multiplying by a beta factor)

<u>Impute Specified Items (aka step 3)</u>:

To streamline processing, we now take the GFAT file and subset the file to get a new file of only those IDs actually needed to be processed further in general imputation processing. This includes those IDs with at least one non blank J<item>00 value or IDs with _COMPLEX='Y'. This file is placed in TEMP##.GFAT. A copy is made to TEMP##.GFATIMP and this GFATIMP file is the input into the imputation program.

1) Ignore IDs on GFATIMP where: BYGIMP = 'B' or that do not meet the users' include clause as defined in file PARMLIB.PSUBSET for the GIMPUTE record. (If the clause for GIMPUTE is blank or status not blank, then only use BYGIMP = 'B').

2) Using the J<item>00 value from GFATIMP, for each ID determine the items items are to be imputed.

3) Ignore ID/items with J<item>00 = 'N'.

4) Impute for sure ID/items with J<item>00 = 'Y'

5) Process each eligible ID using the reject file and values of J<item>00 following the order specified in the

GIS file.

6) For items needing the impbase, determine match key and retrieve and then use in method.

7) If hit a balance complex record, read the GIMP and do corresponding action.

8) As items have been imputed, flag in a manner that they will not be imputed again. (Note however that if you flag an item and later that item is in a complex, completely follow the rules of the complex. If when doing a complex an item is changed, this should be marked and subsequent GIS item entries for that item should now be ignored.)

9) Immediately after <u>each</u> item imputation or balance-complex action, update fat record with revised values of item data.

**WE ARE MAKING THE ASSUMPTION THAT IF AN ITEM IS IMPUTED ON THE ADJUSTED FIELD, ALL REFERENCES TO THAT ITEM WILL ALSO USE THE ADJUSTED VERSION THROUGHOUT IMPUTATION. THAT IS, ONE SPEC_NUM CANNOT USE ADJUSTED VERSION OF AN ITEM AND THEN LATER A SPEC_NUM USES THE EDITED VERSION OF THE SAME ITEM.

Also- put dataflag of M or F (M if the rspcde00 field is I, F otherwise).
Put the impflg (first character of the I<item>pp field) to whatever method code was used to update the item
There are now some occasions where eflg1 will need to be set such as when balance complex doesn't change data and instead flags the total item for review, value of R. Also for items that are to be imputed, but imputation never has valid method, etc., this flag will be set to F. On the fat record, these are found on the first character of the G<item>pp field.

10) NOTE: for year 1, item imputations are RATIO, AUXRAT, VALUE, ATREND, SIMREG.
For year 2, we have added SUM, RESIDUA, PRODUCT.

11) In year 2 we decided to populate eflg2 under certain situations. This was primarily done so that Services could have an indication of administrative data.

For imputed data items, eflg2 is set to the value of the aux1 data flag for the following methods: VALUE, RATIO, AUXRAT, ATREND, SIMPREG.

<u>Update the item file</u>:

1) use compare and update the item file on eddata, impflg, dataflag

2) For surveys that elect to reset item data prior to general imputation runs, the compare program will output a data set that will serve as a transaction data set for future reset operations. Each observation in the data set will represent a changed (imputed) value. The transaction data set will be called DATA00.CMPDETL. PARMLIB.SURVPARM (where parmname='GIMPRESET' and cvalue='Y') indicates that a survey will reset values prior to general imputation.

*if edited fields are different, update the eddata, impflg, dataflg
*else if adjusted fields are different, UNADJUST. See other documentation on how this is done. Then update the eddata, impflg, dataflg

*also check everywhere if any of the 12 char. of G<item>pp is different and if so, update the appropriate eflg# as well.

MIS?????: